Working Paper Series A

Organizational Behavior

Assessing the Behavior and Performance of Teams
in Organizations: The Case of Air Transport Crews

J. Richard Hackman
Robert Helmreich

Number 62

ASSESSING THE BEHAVIOR AND PERFORMANCE
OF TEAMS IN ORGANIZATIONS:
THE CASE OF AIR TRANSPORT CREWS

J. Richard Hackman and Robert L. Helmreich

──────── Abstract ────────

This report examines methodological and conceptual issues in assessing the
behavior and performance effectiveness of work teams in organizations.
The intent is to identify issues of general applicability by focussing
in detail on problems in assessing crews that fly jet transports for
scheduled airlines. Special attention is given to the historical, political,
and organizational context within which assessment takes place, and to
special challenges that arise when teams (rather than individuals) are
assessed.

15 December 1984

ASSESSING THE BEHAVIOR AND PERFORMANCE OF TEAMS IN ORGANIZATIONS:
THE CASE OF AIR TRANSPORT CREWS[1]

About 1815 PST, Flight 173 crashed into a wooded, populated area
killing 8 passengers and 2 crewmembers, and seriously injuring 21
passengers and 2 other crewmembers. The National Transportation
Safety Board determined that the probable cause of the accident was
the failure of the captain to monitor properly the aircraft's fuel
state and to properly respond to the low fuel state and the
crewmember's advisories regarding fuel state. This resulted in fuel
exhaustion to all engines. Contributing to the accident was the
failure of the other two flight crewmembers to fully comprehend the
criticality of the fuel state or to successfully communicate their
concern to the captain.

The Safety Board believes that this accident exemplifies a recurring
problem--a breakdown in cockpit management and teamwork during a
situation involving malfunctions of aircraft systems in flight.

Excerpts from Aircraft Accident Report
NTSB-AAR-79-7

This is one example from a growing body of accident and incident reports
indicating that the functioning of cockpit crews as teams merits further
study. In the above accident, and indeed in most commercial accidents, the
first finding reported from the investigation is that "the flightcrew was
properly certified and qualified for the flight." However, as noted by
Helmreich (1984), recent data from NASA aviation research suggests strongly
that the assumption that technically proficient individuals will form
effective working teams is incorrect. Analyses of safety-related accidents

and incidents show that approximately two thirds of them result from failures in crew coordination (Cooper, White & Lauber, 1979).

Despite increasing awareness of the significance of crew coordination deficiencies in aircraft accidents and incidents, little research has been devoted to the problem. Why is this the case?

For one thing, there is little public pressure for learning more about crew functioning because of the outstanding safety record of commercial aviation. Flying an air carrier is, without question, the safest way to get from one place to another if one examines the number of deaths and injuries per passenger mile travelled. Moreover, even though sub-standard performance by flight crews can result in increased costs (such as fuel and maintenance expenses) and greater-than-necessary risks to safety, poor crew performance usually is invisible to the flying public. Individuals outside the aviation community have little reason to call for additional studies crew behavior and performance.

Another reason for the paucity of research on cockpit crews, one particularly germane to the topic of this paper, is the absence of appropriate methodologies for describing, analyzing, and evaluating cockpit crews. Consider, for example, the accident report excerpted at the beginning of this paper. Several pages of that report are devoted to analyzing a minor mechanical problem which initially distracted the crew's attention. Yet despite the ultimate finding that the crash was due not to the mechanical defect but instead to ineffective crew performance, the analysis of the interaction among members of the crew of Flight 173 is primarily speculative and described in terms of what "could have" or "should have" been done to avoid the crash. There are no generally-accepted methodological tools or

procedures available for assessing how effectively members of a cockpit crew work together.

Objectives and Plan

This paper examines methodological and conceptual issues that arise when one attempts to measure the behavior and performance effectiveness of work groups that operate in organizational settings. We attempt to develop some ideas that have general applicability to team assessment by focussing in detail on one kind of team--crews that fly jet transports for scheduled airlines. We have chosen to focus on such teams for three reasons. First, as will be seen below, challenging issues in assessing team behavior and performance are present in aircraft crews with special clarity and vividness. Second, the stakes are high--assessment outcomes are potentially of life-and-death significance, and both pilots and those who assess them care a great deal about how, and how well, performance assessments are done. And third, we have considerable direct experience with these teams, and believe we can use that experience to frame and discuss some issues that will be of general interest to people who study and manage groups in organizations.

Our aspiration to develop conclusions of general applicability is not without limits, and we begin the paper by conceptually delineating our domain of interest. Then we show how airline crews fit within that domain, and describe how crews function as they go about their work.

We then turn to a discussion of the context within which assessment of airline crews takes place. This is done in considerable detail, because a major point of our paper is that team assessment cannot be done without accommodating substantially to the historical, political, and organizational contexts within which the teams (and their would-be assessors) function.

Then we identify and discuss several special challenges that must be solved by those who would conduct assessments of crew behavior and performance, and we draw on our current research to illustrate some alternative ways to deal with these challenges. While the paper focusses exclusively on cockpit crews in airlines, we hope that readers will find in it some ideas and perspectives that are useful in considering assessment models and practices for a variety of other kinds of teams in other types of organizations.

Domain

Work Teams in Organizations

Our concern in this paper is with the assessment of work teams in organizations. By this we mean teams that are: (a) real groups (that is, intact social systems complete with boundaries and differentiated roles among members), (b) groups that have one or more tasks to perform, resulting in discernible and potentially measurable outcomes of members' collective work, and (c) groups that operate within an organizational context (for more detail regarding specification of the domain, see Hackman, 1983).

This turns out to be a fairly inclusive statement. The domain would include, for example, a group of executives charged with deciding where to locate a new plant, a team of rank-and-file workers assembling a product, a health care team tending to the needs of a group of patients, and a group of economists analyzing the budgetary implications of a proposed new public policy. Nonetheless, many sets of people commonly referred to as "groups" are excluded. Social groups are out (no task), as are reference groups (not an intact social system), coacting groups--i.e., people who may report to the

same manager but who have their own, individual tasks to perform (no group task), and freestanding groups (no organizational context).

Cockpit Crews as Work Groups

Do cockpit crews fall within our domain? Are they real groups, with a real piece of work to accomplish? Or are they, perhaps, mere aggregations of individuals who have their own more-or-less independent work to do in the cockpit, appearing to be a group only because crew members occupy the same small space for a period of time?

Even to raise this question may seem silly: of course cockpit crews are real groups with interdependent work to accomplish. We address the matter explicitly because, as will be seen later, the great majority of existing assessment methods are designed and administered as if success in flying a multi-engine aircraft involves little more than the pre-choreographed execution of individual performances.

Our approach, by contrast, addresses explicitly and in detail the interactive features of cockpit work. So we begin by describing the make-up of cockpit crews and the kind of work they do, to make sure that these groups do fall within our domain of interest.

Crew composition. While the exact composition of cockpit crews varies across airlines and aircraft types, there are enough commonalities among them to permit description of a "typical" airline crew.[2]

There are three roles in the cockpit: captain, first officer (sometimes called "co-pilot"), and second officer (sometimes called "flight engineer"). Pilots move through these roles in a planned, orderly fashion in the course of

---

[2] We will describe a three-person crew, historically the most common size in commercial aviation, although relatively small jet aircraft (such as the DC-9) and advanced aircraft (such as the Boeing 767) are flown by two-person crews.

their careers. A newly-hired pilot begins cockpit work as a second officer. When a vacancy occurs for a first officer position on an appropriate aircraft, the most senior second officer has the opportunity (and, in virtually all airlines, the obligation) to enter a program of training and testing that (if successfully completed) would qualify the individual as a first officer. The pilot serves in that role until reaching the top of the first officer seniority list, at which time he or she begins another program of training and testing to qualify for a captaincy.

Duties are clearly defined for each role. The captain has overall responsibility for the flight and for management of the cockpit crew. The captain cannot be ordered to undertake a flight by airline management (or by anyone else) if in his or her judgment the flight would be unsafe (e.g., because of mechanical or weather problems). The first officer shares flying duties with the captain, and normally flies every other leg of a trip. The captain can take control of the aircraft at any time--for example, in particularly challenging circumstances. If the captain is flying, Federal Aviation Regulations allow the first officer to take control only when he or she observes that the captain is incapacitated (e.g., ill or severely emotionally distraught). But it is professionally risky for a first officer to do this, and it happens very rarely. The second officer controls the mechanical systems of the aircraft (the engines, fuel, the electrical and hydraulic systems, and so on). He or she conducts the external walk-around inspection of the aircraft before each departure, and is the primary point of interface with the cabin crew (for example, adjusting the air conditioning or attempting to repair cabin equipment that malfunctions in flight).

Individual crew members bid for sets of flights (called "trips"), and in most airlines requests are honored in order of seniority. The composition of a given crew, then, depends both on the bids submitted by its members and the assignment rules used by the airline's crew scheduling system. Crew members typically are rostered together for one month (the usual airline bid cycle), but it is not uncommon for their time together to be shortened or interrupted because of vacations, training schedules, or personal matters. Some pilots bid for (or may be assigned) "reserve" duty, filling in for absent crew members as needed.

Work activities. Crew members meet for the first time in the airline's flight operations office (or, occasionally, in the cockpit). They may or may not have a structured briefing about the trip to be flown, depending on the airline's policies and the captain's proclivities. A day's flying may involve a single long flight (e.g., transcontinental) or as many as half a dozen short segments. At day's end, the crew may wind up at members' home base (in which case individuals are likely to head for their personal homes as soon as possible) or at a distant airport (in which case crewmembers are likely to spend considerable time together in social or recreational activities).

The actual tasks performed are of five general types: (a) planning and decision-making, including reviewing flight plans, making operational decisions in flight, and dealing with abnormal circumstances; (b) manipulating the flight controls (i.e., actually flying the airplane), (c) monitoring and adjusting various mechanical and electrical systems, such as navigational equipment and the aircraft's engines; (d) completing paperwork, such as computing the "weight and balance" form prior to departure, and entering various data in logbooks, and (e) communicating with other individuals and

groups who are involved in the flight (specifically, air traffic controllers, the airline's flight operations and maintenance staffs, and the cabin crew on board the aircraft).

The crew's workload is very uneven, and typically is bimodal--with substantial work on all five types of tasks occurring near the beginning of a flight (preparing for departure, take-off, and climb) and then again near the end (planning the approach to the destination airport, executing the approach and landing, and "closing the books" upon arrival at the gate). During these two periods, all three crewmembers are quite busy, and a great deal of communication and coordination among them is required. If an unusual situation develops during one of these periods, the capacity of the crew can be pushed to its practical limit--posing a considerable challenge to the captain's leadership skills and the capability of members to function as a team. During the time that the aircraft is cruising at its assigned altitude, on the other hand, performance demands are minimal. Indeed, on long and uneventful trips, crews often have to work hard to fend off boredom during the cruise portion of the flight.

Summary. Do cockpit crews fall within our domain? Are they intact social systems, even though they are small in size and have a relatively short life span? Yes. Do they have a set of tasks to be performed whose outcomes can be discerned and, potentially, assessed? Yes. And do they operate in an organizational context? Yes--many contexts. Cockpit crews, for all their unique features, clearly qualify as organizational work teams.

Yet the uniqueness of these teams must not be overlooked, because the special features of cockpit crews pose some major challenges for those who would assess them. The teams are, for example, both temporary and composed of

individuals who typically did not choose to work together (assignments having been made by a computer in response to <u>individual</u> bids and seniority). Moreover, team members usually have little time to get to know one another before their first period of demanding collaborative work begins. Also noteworthy is the variance in workload: long periods of routine activity, punctuated by demands for intense and highly interdependent teamwork--some of which are predictable ahead of time (such as landing in marginal weather), but some of which are not (such as an extended and unexpected hold or wind shift that raises questions about the sufficiency of the fuel on board).

The Context of Cockpit Crew Assessment

Let us now turn to an examination of the context within which the assessment of cockpit crews takes place--for it is this context that shapes both what is <u>appropriate</u> and what is <u>feasible</u> in designing, conducting, and using the results of a team assessment program.

Current Practice

Federal Aviation Regulations require pilots to be assessed on a regular basis. These assessments include a "proficiency check" and/or a "line check." The line check consists of observations of the pilot's performance on a regularly scheduled flight. The proficiency check involves flying a series of required maneuvers in an aircraft simulator. These maneuvers address both technical skills and emergency procedures, such as steep turns, loss of an engine, aborted take-offs, landings with an engine out, missed approaches, and precision and non-precision approaches.

The frequency of checks required varies as a function of position (captains are evaluated more frequently than first officers or second

officers). The evaluation may legally be conducted either by an FAA inspector or by a Check Airman, a pilot designated by the air carrier and approved as an evaluator by the FAA. Whether the evaluator is from the FAA or is a Check Airman, the only possible outcomes of a check are "pass" or "fail." A pilot who fails is re-examined after additional training. Failing the re-examination results in loss of license and, hence, loss of the right to function as a crewmember in commercial airline operations.

Anecdotal reports from FAA officials, Check Airmen, and other airline officials, as well as the personal observations of the authors, support a view that this dichotomous classification of acceptability as a flightcrew member masks a wide range of performance variability. Moreover, the focus of evaluation in the proficiency check is a pilot's ability to demonstrate individual technical proficiency in the control of the aircraft under a standardized set of conditions. What is distinctly not measured in this evaluation is the pilot's ability to evaluate alternatives and make decisions in a complex, stressful environment, to draw appropriately on the knowledge and perspectives of coworkers, and to coordinate one's own work activities with those being performed by other crewmembers.

These omissions are particularly worrisome for captains, whose role requires them to manage a complex array of technical and human resources, and to employ those resources effectively in non-standard situations. A significant proportion of accident analyses implicate poor leadership and management as causal factors. Typical is a case in which a captain fails to respond to input from crewmembers indicating that the captain's behavior is endangering the flight. Recall, for example, the incident referred to in the opening paragraph of this paper. The captain disregarded repeated warnings

that the fuel state was dangerously low while preoccupied with the possibility that the landing gear was not locked in the down position--and the aircraft eventually ran completely out of fuel and crashed.

In general, only pilots who are obviously and dangerously incompetent fail checks, and even they have a high likelihood of passing upon re-examination. It is not possible (because of the simple pass-fail criterion used) to estimate how much variation there is among those who pass their checks. Nor is it possible to determine with existing data whether or not existing check procedures address those aspects of performance that are most critical to flying as a member of a two- or three-person cockpit crew.

Historical Context

Psychologists interested in assessment have been involved with aircraft crews for several decades. During World War II, for example, American psychologists were mobilized to help solve the practical problems surrounding the selection and training of large numbers of military pilots.[3] Throughout the war, the criterion used in selection research was completion of (vs. elimination from) pilot training. The investigators were plagued by the fact that this criterion was largely subjective. Although attempts were made to standardize grading and to obtain ratings from multiple instructors, subjectivity in evaluator judgment was not eliminated.

Forty years later, subjectivity remains a disconcerting issue for both pilots and their evaluators. While criteria for standard evaluations have improved and computers allow the precise measurement of how flight controls

---

[3] Because of the urgency and importance of the air war, some of the most outstanding talent in psychology was applied to pilot selection problems. Much of the research accomplished was compiled and edited by Arthur W. Melton after the war (Melton, 1947). This volume shows the origins of many of today's practices and illustrates the continuity of many problems in pilot evaluation.

are manipulated in aircraft and simulators, the critical areas of judgment, leadership, and decision-making are still rated subjectively. There have been few attempts to train evaluators in how to assess these "soft" aspects of pilot competence, or to develop standardized ways of measuring them.

In one of the major studies of training success conducted in 1942, the relative importance of four major categories of performance was tabulated by computing the percentages of pilots eliminated from training who had been cited as deficient in each (Melton, 1947). The categories were: (a) coordination and technique, (b) alertness and observation, (c) intelligence and judgment, and (d) personality and temperament. Results showed that 81 percent of the failures had to do with poor coordination and technique--with the consequence that subsequent training and evaluation programs placed by far the greatest emphasis on the technical, "stick and rudder" aspects of flying.

Although intelligence testing was (and is) included in most pilot selection programs, personality factors have received relatively little attention. When personality assessment is employed, its use has been primarily to screen out individuals on the basis of actual or potential psychopathology. Few efforts have been devoted to selecting in individuals on the basis of personality attributes associated with particularly effective performance--e.g., by identifying characteristics associated with pilot effectiveness and using these characteristics to select individuals from a pool of technically qualified applicants.

The concentration on individual proficiency rather than crew effectiveness, a hallmark of current assessment practice, also has its roots in history. The tradition in the military has been to give individuals at the top of their classes first choice of aircraft type. Most choose single pilot,

fighter aircraft, leaving multi-pilot bombers and transports to their less proficient colleagues. Given the coordination and agility required for single-engine combat in World War I, and the white scarf tradition of the Red Baron and Captain Eddie Rickenbacker, this philosophy was probably justified. Today, given the different skills and aptitudes required to fly a complex multi-engine jet aircraft in a crowded and demanding air traffic environment, it probably is not. Yet, as seen in the previous section, airline pilots continue to be evaluated as individuals, and are assigned grades of "pass" or "fail" based mainly on their skill in manipulating flight controls.

Perspectives and Stake of the Airlines

It is clearly in the interest of airlines for cockpit crews to perform as competently as possible. A crash, for example, has severe financial consequences for the company--beyond the incalculable personal costs to those involved. Revenue is lost because potential passengers avoid the carrier, insurance rates (a major cost item for the airlines) rise, and investors may develop second thoughts about the wisdom of owning the airline's stock. Good performance in the cockpit also contributes directly to an airline's financial well-being. On-time performance may be improved (which can result in a reputation for reliability that attracts passengers), the amount of fuel burned on a flight (another major expense) can be significantly reduced, and maintenance delays and costs can be minimized.

Yet despite the demonstrable benefits of improving flightcrew performance, U.S. airlines have been notably non-aggressive in seeking more comprehensive evaluation of flight behavior and in striving for higher levels of crew performance. Many airline executives may feel that the economic challenges they face (which are of obvious relevance to long-term corporate

survival) take precedence over the pursuit of improved crew effectiveness--a not unreasonable position, given the overall safety record of the industry. There are, moreover, some seemingly good reasons for executives not to push for broader and more intensive assessment of cockpit crews. One has to do with the impact of deregulation on corporate priorities, one with the state of labor relations in the industry, and one with the legal risks of maintaining records that document variations in pilot competence and performance.

The impact of regulation and deregulation. Until 1978, both the routes flown by individual carriers and the fares charged were controlled by the Civil Aeronautics Board. During this period, carriers were given generally non-competitive assignment of routes, and passenger fares were federally controlled to provide a "reasonable rate of return" to the airlines--even including subsidies for carriers flying to certain destinations where traffic was light. There was little incentive to contain costs since they could be passed on to passengers with federal blessing.

After deregulation of the industry in 1978, airlines found themselves in a fully competitive environment where routes were freely available and where fares and profits would be determined by the free play of the marketplace. Predictably, this resulted in greater attention to costs, and programs that could not be shown to contribute directly to an airline's ability to compete often were eliminated or reduced in size.

Investments in research and development for pilot training and evaluation were substantially reduced by many airlines--and just at a time when flight training staffs were beginning to recognize that crew dynamics were critical to the safety of flight. Moreover, the increased competitiveness of the airline industry appears to have lessened the sharing that traditionally had

characterized relations among flight training groups in different companies. The net result was that individual airlines had less material relevant to crew training and assessment to share and less incentive to share it than they had prior to deregulation. The Federal Aviation Administration (FAA), which might have picked up the research and development activities being curtailed by the airlines, did not do so.

Performance evaluation and labor relations. U.S. pilots and their professional (union) organizations generally have opposed increases in formal pilot evaluation (for reasons to be explored below). In recent years, the airlines have had little incentive to press the issue. In the early 1980s, established carriers felt the double jeopardy of an economic downturn (which reduced loads and revenues) and intense competition from new, low cost carriers with nonunion workforces. In response, a number of airlines asked for significant concessions in wages and work rules from pilots. These negotiations have been delicate and important, and most airlines have avoided or deferred any issue that might turn them sour. It is, then, not surprising that there has been little pressure from the established airlines to increase the scope or intensity of pilot evaluation.

The newer, low cost airlines, on the other hand, having already obtained a pilot force willing to work longer hours and undertake more varied responsibilities for less pay, were not motivated to upset this profitable and productive state by imposing performance evaluation standards more rigorous than those of the established carriers. As a consequence, virtually all carriers have stayed clear of evaluation issues and have simply complied with federally mandated standards.

Potential for liability. An airline that collected and maintained assessment data documenting differences in pilot competence and performance could be especially vulnerable to litigation in the event of an accident. If, for example, an accident were found to be caused by "pilot error" and if it were further determined that assessment data for crewmembers on that flight placed them below the carrier's average, then litigants could argue that the airline had callously endangered passengers' lives by boarding them on a flight staffed by substandard personnel. A case in point is the Air Florida jet that crashed into a bridge shortly after takeoff from Washington National Airport (NTSB, 1982). Pilot judgment and performance were determined to have been causal factors in that crash--and it happened that the captain had failed a proficiency check prior to the accident (although he had passed the examination after retraining). It is not possible to determine the precise effect this disclosure had on the outcomes of lawsuits and the subsequent failure of the airline, but its impact was clearly negative.

Perspective and Stake of Pilots

U.S. pilots have generally opposed changes of current performance evaluation practices. Moreover, they have resisted proposals to increase the quality and scope of data obtained from Cockpit Voice Recorders and to make data from Flight Data Recorders accessible to aviation researchers. Organized opposition has been spearheaded by the Air Line Pilots Association, the largest and most powerful union representing airline flight crews.

There are conflicting interests for both pilots and their representative organizations. Obviously, it is in pilots' personal and professional interest to achieve a high degree of safety and to promote the financial health of their employers by enhancing operational efficiency. On the other hand,

negative performance evaluations can result in loss of license and professional livelihood.

At first glance, it might appear that pilot opposition to comprehensive performance assessment represents a triumph of narrow self-interest over a collective good. Many of pilots' concerns about how assessment data are collected and used are, however, well founded. Subjectivity in evaluations, for example, has been and continues to be a real problem. The recent emphasis on assessing the decision-making and managerial skills of captains (and the capability of the crew as a whole to work together effectively) has increased the salience of concerns about subjectivity. To date, the technology of evaluation and the training of assessors have not advanced far enough to reassure pilots that evaluations of the non-technical aspects of their performance will be accomplished reliably, validly, and impartially.

Adding to the evaluation anxieties of pilots is the fact that labor-management relations between pilots and airlines have been more adversarial than collegial in recent years. Part of this conflict grew from the fact that pilots typically earned significantly more money for significantly less time at work than non-flying middle and upper managers. While this situation has been changing dramatically since deregulation, there is still a perception among pilots that management would like to use evaluation as a club to bring pilots into line. It would be possible, for example, to use subjective evaluations to terminate individuals who are particularly effective spokesmen for pilot concerns; or, perhaps, cockpit voice recordings of flights flown by these individuals could be subjected to special scrutiny as a means of discouraging dissent.

Finally, pilots (like their managements and federal regulators) tend to perceive the crew as an aggregate of individuals rather than as a team with the captain as manager/leader. Helmreich (in press) found that 66 percent of captains agree with a statement that command performance is not adversely affected by having an inexperienced or less capable crewmember in the cockpit, while 92% believe that they should take control and physically fly the aircraft during nonstandard or emergency situations. Many first officers' attitudes fit well with this view: 29% of those surveyed state that they should not question the decisions or actions of the captain except when there is a direct threat to the safety of flight.

In sum, the reluctance of pilots to endorse changes that would expand the scope or intensity of performance assessment is quite understandable--for reasons of self-interest, certainly, but also because of problems with the quality of the tools available for collecting data about the non-technical aspects of pilot and crew performance.

Perspective and Stake of Federal Agencies[4]

The FAA is charged with mandating practices that will ensure the highest level of safety in commercial aviation. However, the FAA must also respond to a number of conflicting pressures. While safety is presumably paramount to the FAA, the agency also recognizes the need to promote civil air transport and is sensitive to pleas from carriers regarding the financial consequences of proposed regulations. Moreover, the FAA is subject to direct lobbying activity--both from representatives of pilots' organizations (who may argue

---

[4] We discuss here only the FAA and the National Transportation Safety Board (NTSB). While these are the primary agencies directly involved with crew performance and flight safety, it should be noted that NASA also contributes to these issues by conducting research on aeronautical topics, and by advising both the airlines and the FAA.

that their constituents would be harmed by certain regulations) and from
passenger and public interest groups (who often seek more stringent controls
on pilot behavior and more thorough evaluations of crew competence and
performance).

The strongest advocate of improved performance measurement and evaluation
is the National Transportation Safety Board (NTSB), a federal agency charged
with determining the causes of accidents and recommending procedures to avoid
their recurrence. Based on its analyses of data from a number of airline
crashes, the NTSB has repeatedly recommended that the FAA increase
requirements for data capture by Flight Data Recorders and Cockpit Voice
Recorders, and that greater emphasis be placed on training in assertiveness
for junior crewmembers and in crew coordination for all pilots. Despite the
weight of the NTSB data and recommendations, the FAA has been slow to change
regulations governing pilot training and assessment. Given the political
forces to which the FAA is subjected, it is doubtful that the organization
will become significantly more aggressive in these areas in the forseeable
future.

Summary

This section has laid out some of the factors that impede innovation in
the assessment of flight crews as task-performing teams. The list is long: a
strong historical emphasis on assessing pilots as individuals on a pass-fail
basis, cost considerations that are increasingly important to airlines in a
deregulated competitive environment, the felt need by pilots and their unions
for protection from biased evaluations and disciplinary actions, the
deteriorating labor relations climate in the airline industry, airlines'
concerns about their liability for the results of accidents, and even the

uncertain relationship between the two major federal bodies concerned with aviation (the FAA and the NTSB).

Even if one had a superb, validated method for assessing the behavior and effectiveness of crews qua crews, one could not simply present it to the airline community and expect it to be adopted. There are, for example, technologies already available that could be used to improve crew assessment and training, such as multi-channel digital flight data recorders used for operational analysis in Europe and found valuable there. Yet these devices are found only on wide-body aircraft in the U.S., and then only because they were installed by the manufacturer when the planes were built.[5]

Whatever new procedures or devices are devised for assessing cockpit crews, they must be be adopted and used within the relatively constraining historical, political, and organizational context described above. Contextual factors, too often overlooked by psychologists charged with the design of psychometrically sound assessment devices and procedures, strongly condition what one can do, and what one can reasonably expect to accomplish, in assessing cockpit crews in U.S. air carriers.

Challenges in Assessing Cockpit Teams

Having explored the context within which assessment of cockpit crews takes place, we now identify and discuss several challenges in the actual conduct of such assessments. As will be seen, a number of opportunities to obtain particularly informative data lurk just behind the challenges described below.

---

[5] If a wide-body aircraft should crash, NTSB investigators would have the benefit of data provided by one of these recorders. These exceptionally informative data cannot be used for crew training or assessment, however, even though they are automatically collected during every flight of these planes.

1. <u>A great deal of assessment and regulation of flying performance does occur</u>
   <u>in airline organizations--but the form of those activities make it of</u>
   <u>limited use to organizational representatives responsible for proficient,</u>
   <u>safe flying operations.</u>

Pilots constantly assess one other--although they would not use that word. Airline flight operations departments buzz with conversation about flying and about pilots. This is understandable, given that people generally like to talk about their work. Pilots seem especially fond of talking about who is a great pilot, who is shaky, and who is and is not a good team player in the cockpit. While these conversations are, in some ways, like the gossip one hears in the coffee rooms of any organization, they are more than that: pilots are talking about things that are potentially life- or license-threatening. For all the humor that characterizes such discussions, pilots <u>care</u> about what is being said, and they store much of it away for future reference.

The focus of the informal assessments pilots do is on individuals, not crews. While there are plenty of stories exchanged of the type "So there <u>we</u> were at 35,000 feet..." the assessments and attributions that are made are almost invariably about individual crew members. One might, for example, hear something like this:

> "...so there was this flock of geese having a tea party right over
> 22 Left [a runway designation], and the tower switched them to 29
> just when Charlie was getting lined up on the ILS [instrument
> landing system]. Well, the weather was a mess, they were vectoring
> old Charlie all over the place, and he got confused and got behind.
> Three times Phil had to remind him about something, and eventually
> Phil just took it and landed the damn thing."

One would be far less likely to hear an account of the same set of events that went like this:

"...so after they got ATIS [a recorded radio transmission giving weather and runway information] they just assumed it would be a routine ILS approach to 22 Left and they started chewing the fat. They didn't hear the talk on the radio about the geese over the runway, so when the tower switched runways at the last minute it was scramble time. Charlie was flying, and he had his hands full because of weather and the new vectors he was getting. Phil started changing the radios to set up for the new approach, but didn't tell Charlie what he was doing--and Charlie couldn't figure out what the hell was going on. Nobody really got things organized, everybody got confused, and eventually Phil got so frustrated that he took the airplane and landed the damn thing himself."

In the first account, the one most likely to be heard, Charlie has a problem--he let a situation that was not all that demanding get the better of him, and he had to be bailed out by Phil, his captain. The attributions made are all to individuals. The second account invites a group-level interpretation: the crew got itself into trouble, by not paying attention to changing situational demands, by not planning and organizing the work (either contingently beforehand, or in real time after the runway change was announced), and by poor between-member communication and coordination. Indeed, if someone is to be blamed in this situation, it might most appropriately be Phil for not managing his cockpit well--an interpretation unlikely to be made based on the first account, in which Phil is implicitly viewed as the savior.

This illustration is not meant to imply that most attributions of responsibility for negative events are made to junior crew members. Indeed, the opposite is more often the case: There is rich lore in every airline we know specifying which captains have what quirks. People talk incessantly about the personality and behaviors of their leaders, and captains are not exempt from such talk. The point, instead, is the individualistic orientation of the informal assessments made by airline pilots. This is not surprising,

given the focus of airline selection, training, and evaluation programs. But it does suggest that most pilots may be neither experienced nor comfortable making group-level assessments and interpretations about what happens in a cockpit--even though, as in the example given above, it often is the crew, as a crew, that gets itself into trouble.

The informal assessments pilots make of one another do result in some informal regulation and pilot-to-pilot coaching and counselling. At the extreme, certain captains are known to "run a bad cockpit," and are not to be flown with if at all possible (even, in some cases, to the extent of calling in sick if one is rostered with that captain). More gentle are data about how a crew member needs to behave with some captain (e.g., "don't make any suggestions, he bristles if you do"), or advice about help a given crewmember needs (e.g., one captain telling another about the particular flying foibles of a first officer). These data are in the system, but they are not available to the system--and certainly not to the regulatory aspects of the system (i.e., the FAA, check airmen, or airline managers). Pilots, for all their concern with safety, are also members of a fraternity: one protects another from potential disciplinary action, with the confident expectation that the reverse will be true should the tables someday be turned.

In sum, there are rich assessment data already available in every airline, and those data are used to some extent for self-regulation by the pilot community. But the data are kept strictly within that community, and they mainly have to do with the behavior and skills of individuals. The potential of informal assessment data for pilots' learning (about themselves as individuals, and about their functioning as teams) is considerable--for example, through a systematic program of peer feedback and group self-

assessment. Given the political and organizational realities discussed earlier, however, it will not be easy to find ways of using these data systematically to foster pilot and crew effectiveness.

2. Objective indicators of crew performance are incomplete and inadequate--perhaps inherently so.

It is common, when discussing strategies for assessing task-performing teams, to call for collection of "objective" performance measures. There are three reasons why we do not join in that call.

First, truly significant hard data (i.e., the occurrence of a crash or serious incident) become available very infrequently. Therefore, these events are useful mainly in retrospective analyses of the technical and human factors that may have contributed to them. The NTSB conducts these investigations, drawing on a variety of data (including those from Cockpit Voice Recorders and Flight Data Recorders), and much is learned from them. But, fortunately, there are few occasions to conduct them and for that reason they do not play a major role in the day-to-day assessment of airline pilots and crews.

Second, the completeness and quality of available hard data are quite limited. Flight Data Recorders on the majority of aircraft in service in the U.S. provide only for analog recording of limited data on metal foil, and Cockpit Voice Recorders yield low-quality recordings (from a single cockpit microphone) on a continuous-loop thirty minute tape. Even these relatively primitive data cannot be used except by the NTSB in the case of a reportable accident or incident--in contrast with practice in Britain, where multi-channel digital data are collected for every scheduled flight and used both to develop statistical summaries and to counsel individual pilots (for a more complete description of British practices, see Helmreich & Hackman, 1984, and

Mearns, 1983). Again, political realities make it doubtful that more sophisticated and complete "hard" data will be available for use in crew assessment in the near future. Even in Britain, labor-management agreements require that pilots' identities be kept confidential (except in the case of serious or repeated lapses from safe practice), which limits the usefulness of the data for assessment purposes.

Finally, even if data from Flight Data Recorders were more complete, of higher quality, and more readily accessible to assessors (whether airline personnel concerned with flight standards or researchers) they would be of limited use for crew assessment. For one thing, these data address only technical, "stick and rudder" issues--and, moreover, they serve mainly to identify bad performance, such as control manipulations that lie outside acceptable parameters, or deviations from correct procedures or flightpaths. More importantly, hard data provide no clues about how well the crew, as a task-performing team, has functioned. Even the British measures, which are probably the best presently available, are not analyzed (and, by their nature, probably cannot be) in a way that would allow assessment of cockpit resource management and crew coordination issues.

The problem with objective performance measures is, at root, conceptual rather than technical or methodological. Just as there are multiple routes one can fly and still get from New York to Chicago, so are there multiple ways that a crew can operate and still achieve essentially the same performance outcome. Systems theorists (e.g., Katz & Kahn, 1978) call this property of social systems "equifinality," and it is one reason why simply looking at a given outcome (e.g., arriving safely in Chicago) may not tell one much about how well the cockpit crew functioned. The phenomenon of equifinality

obviously complexifies the assessment task, as does Tyler's (1983) notion of "multiple possibilities." Tyler asserts that there are many possible outcomes that can emerge in any given situation, and the particular one actualized is not completely determined by the causal factors that precede it. Multiple possibility theory envisions a world with some "play" in the system, and it encourages attention to human and social choice as a factor that transforms multiple possibilities into single courses of action.

So where equifinality alerts us to the fact that the same outcome can occur in response to many different causes, multiple possibility theory posits that the same cause can generate a variety of different outcomes. Taken together, the two notions call into question assessment methods that assume that single causes (e.g., certain behaviors in the cockpit) are tightly linked to specific performance outcomes (e.g., optimally efficient fuel burn--one of the measures that could be obtained from a sophisticated Flight Data Recorder).

In sum, while it would be good if more and better hard data were available, the likelihood of that happening in the existing organizational and political context is low. Moreover, even if such data were available they would be of limited use in crew assessment because the link between how members of a team behave and eventual group performance outcomes is not a tightly-coupled, deterministic relationship in which specific behaviors always lead to a given performance outcome. Objective performance data simply do not provide a sturdy or complete enough base on which to build an robust cockpit crew assessment program.

3. <u>"Process criteria" of performance provide an alternative to objective</u>

   <u>measures--one fraught with both risk and opportunity.</u>

If hard outcome measures are not obtainable (or fully appropriate) for use in assessing cockpit crew performance, can observations of the performance <u>process</u> of crews as they work be used instead? In fact, this is already being done, and with success, for certain kinds of performance situations--which we will call, for want of a better term, "acute" situations.

Since crews are rostered temporarily (and therefore do not have time to develop their own strategies for handling all situations they might encounter), airlines have developed highly standardized procedures to be followed in unusual or particularly demanding circumstances. One example is a "Category II approach," in which the crew lands an appropriately instrumented airplane on instruments in low visibility conditions. A Category II approach requires extremely close coordination among crew members at a critical time (i.e., the instant when a decision must be made either to land or to execute a missed approach). Other acute situations include an engine fire warning, instructions from Air Traffic Control to change course immediately to avoid a collision, and so on. In each of these cases, all crew members are trained beyond proficiency in their specific duties, and when the triggering event occurs, the prescribed processes are executed precisely as previously choreographed and practiced. A crew of well-trained strangers should be able to handle an acute situation just as competently as a crew that has flown together for many weeks.

Because there is only one right way to behave in most acute situations, it is reasonably straightforward for an assessor (one who is expert in the procedure, of course) to determine how well the team handled it--and, if

mistakes were made, to specify exactly what they were and who made them. Process criteria provide an appropriate way to assess crews in acute situations, and check airmen routinely use them in simulator exercises to help crew members become proficient in performing their parts of overall team task.

The use of process criteria to assess cockpit crews is a very different undertaking when the situation is not acute. In these circumstances, which we will call "continuing situations," conditions require a decision making process involving consideration of alternative courses of action and the development of a shared strategy for action. These are situations which are not overlearned and where only general training and experience is relevant. Examples include mechanical malfunctions that do not pose an instantaneous threat but place in jeopardy the safe continuation or completion of a flight (e.g., landing gear problems, or engine, hydraulic, or electrical difficulties). These are problems that require the coordinated action of the full crew and, not surprisingly, are the kinds of situations frequently encountered in incidents and accidents where conclusions of "pilot error" are reached.

It is more difficult to use process criteria of effectiveness in continuing situations. There are, to be sure, better and worse ways to handle them, and how a given problem is dealt with can significantly affect both the likelihood that new problems will develop later, and the capability of crew members to work together competently later in the flight. Yet these "better and worse ways" cannot be specified in advance the way one can for an acute problem, and that makes the assessment task considerably more challenging.

Competent check airmen report that they are able to sense how well a crew handles continuing problems. And, after a period of time observing a crew,

they may confidently conclude that Captain X is a "poor leader" or that members of a given crew "have real problems working together," although they often are unable to articulate the precise reasons for these judgments. When pressed for evidence, check airmen tend to talk about poor decision-making processes, slippage in coordination among crew members, and incomplete or inadequate communication--rather than about the technical aspects of flying.

Such talk makes them uncomfortable, even though they invariably discover, when they check with their colleagues, that others have very similar assessments of a given pilot or crew. The discomfort is strong enough that a number of check airmen have expressed to us real doubts about whether such "soft and groupy" matters are legitimate for them to address at all. These items, they say, are wholly ignored by the FAA in its requirements for pilot assessment--so why should we take them so seriously? But they take them seriously nonetheless, partly because the FAA does focus so exclusively on individual technical proficiency. Assessments of leadership and team processes in the cockpit, for all their subjectivity, fill an important void.

If check airmen are to become more comfortable, and more competent, in assessing cockpit crews as teams, they will need both (a) tools for doing so, and (b) training in the appropriate use of those tools--neither of which is presently available. Development of such materials is, in our view, work well worth doing, and we will have more to say about it (including discussion of a technology that may facilitate that work) below.

4. Many events important to competent crew functioning occur outside the cockpit.

It is natural to look where the key team behaviors are actually occurring if one is interested in assessing how well a team is functioning. In our

case, that obviously is the cockpit. But there are problems with focussing
exclusively on what happens in the cockpit.

First, while the cockpit is where the team does its work, the crew
typically is formed and disbands (for the day, or permanently) elsewhere.
What happens in the flight operations office (where crews check in, get their
dispatch releases, and perhaps have a cup of coffee) can be critical to team
functioning, especially at the moment when crewmembers meet and form their
first impression of the captain. Similarly, what happens at the end of a
day's flight, perhaps on the crew bus or over dinner, can have a profound
influence on subsequent crew performance (at one extreme, dinner can serve as
an extended de-briefing session that strengthens the team as a performing
unit; at the other, it can strain relationships among members in a way that
damages their ability to work together subsequently). We know from group
research (e.g., Gersick, 1983; 1984) that the beginnings of groups, their
midpoints (such as the evening at an outstation on a two-day trip), and their
endings are especially critical in understanding a team. It would seem
advantageous, therefore, to address these non-cockpit times in assessment
methodologies.[6]

Second, there is increasing recognition of the importance of the
organizational context in determining how groups function. Organizational
features such as information systems, reward practices, control procedures,
available communication channels, and even the way physical space is designed
have significant effects on crew behavior and performance (Hackman, 1983).

---

[6] The NTSB has begun to collect and analyze data of this kind in its
investigations of accidents. In analyzing the 1981 crash of a Cascade
Airways Beach 99A, for example, the NTSB explored in detail both how the
crew functioned on previous legs of the fatal flight, and recent events in
the personal lives of crew members (NTSB, 1981).

Consider, for example, an airline that had a driving commitment to on-time performance, with bonuses for crews that consistently achieved company targets. Such a reward system surely would alter crew dynamics, and might even tempt crews to take shortcuts that could waste fuel and/or compromise safety.[7] Or consider what can happen to a crew in an airline where there are too few operations personnel available to handle all the radio requests received from cockpits on bad weather days. A crew observed by one of us discovered, in flight and in rapid succession, that (a) the airport from which it had just departed had closed, (b) its destination airport had closed, (c) weather at its alternate airport was deteriorating fast and that airport was expected to close, and (d) it was not possible to get the attention of a dispatcher (because all dispatchers were already fully occupied with other urgent business). At that point, the captain became extremely autocratic and evaluative in his dealings with other crew members (behaviors he had not exhibited previously in the flight), and the climate in the cockpit became tense and sullen--a climate unlikely to foster effective team problem solving and decision making. Finally, consider something as mundane as the existence of a quiet briefing room, where pilots can get psychologically prepared for their flight. The simple presence or absence of such a facility can have strong effects on how crew members relate to one another when they first start their work together. And those first encounters, in turn, can establish a style of interaction that may be difficult to change for the rest of the day--or the rest of the month.

---

[7] The "fast buck" program initiated by Braniff International in 1968 required the airline to pay each passenger a dollar if a flight did not arrive at its destination within fifteen minutes of schedule. This program may have contributed to the crash of a Braniff Electra turboprop in May of that year. The flight had been delayed on departure and was pushing the fifteen minute limit as it neared the destination airport. The crew attempted to penetrate a line of thunderstorms rather than navigate around them, and lost control of the aircraft in turbulence (Nance, 1984, Ch. 6).

If it is true that structural and contextual factors condition crew interaction (and we believe the evidence is clear that they do), then any robust assessment methodology should include measurement of such features. Without such data, it may not be possible to correctly interpret what is observed in the cockpit. Moreover, it may be that interventions intended to correct poor team behavior should focus on the larger organization in which the crew operates rather than on specific exchanges that take place among crew members. Assessors of cockpit crews must be alert to organizational influences, and not fall into the trap (a trap already well-populated with disheartened small group researchers) of acting as if member interaction is all that needs to be examined if one wishes to understand and evaluate a task-performing team.

5. An assessment system that is appropriate for determining training needs can be inappropriate for the evaluation of crewmembers--and vice versa.

A classic issue in organizational performance appraisal is the tension between using assessments for training and development purposes vs. for evaluation and control (see, for example, Porter, Lawler & Hackman, 1975, Ch. 11). Training-oriented assessments, while they may be anxiety-arousing for the assessee, are consequential mainly for his or her own learning and development. Evaluation-oriented assessments, on the other hand, are more broadly consequential and may, for example, affect the size of one's raise, the probability of a promotion, or even the security of one's job.

Organizations, understandably, want to use appraisals for both purposes, and many managers have sought assessment techniques that can be used simultaneously for training and for evaluation--procedures that provide

incentives for people to learn while discouraging them from "gaming" the process to secure a favorable outcome. Such procedures are hard to find. Even to search for them can be risky, in that attempting to achieve both objectives can sometimes result in achieving neither.

Although the trade-off between training and evaluation is relevant to all aspects of crew assessment, the tensions are especially vivid in Line Oriented Flight Training (LOFT)--a program that is arguably the most significant development in aircrew training in recent years. In a LOFT exercise, a complete two or three person crew undergoes the simulation of an entire line flight between cities. The goal of the simulation is to reproduce the complete flight environment including dispatch releases, weight and balance computations, en-route weather, and communications with the cabin crew, Air Traffic Control, and company operations. Typically, one or more abnormal or emergency situations are introduced during the flight. Aviation psychologists, especially those associated with NASA, have been heavily involved in the development of LOFT and have developed guidelines for maximizing the training benefits of the experience (Lauber & Foushee, 1981).

Even highly experienced crews report that LOFT is a powerful training tool that allows them to test all their skills, both technical and managerial, under extraordinarily realistic conditions. Crews can gain many valuable insights from the experience itself, especially when the simulation is videotaped and can be reviewed by the full crew, and when the debriefing is conducted by a competent and credible trainer. When meaningful measures of team processes and outcomes become available (a matter for which we intend our own research to be helpful) the power of LOFT technology for individual and team training should increase even more.

Although originally conceptualized as a training tool, LOFT also is useful for formal evaluations of pilot competence. It is relatively straightforward, for example, to construct scenarios that allow observation of performance on complex but standardized flying tasks; in addition, special scenarios can be developed that allow observation and assessment of behaviors that may be of concern for a certain pilot.[8] The FAA has recognized the usefulness of LOFT for evaluation, and has approved the substitution of a LOFT exercise for one of the annual checks required of all pilots. In doing so, the FAA also instituted a requirement that performance must be "satisfactory"--i.e. it must meet the general standards applied in evaluating individual pilots in a simulator or line check.

This requirement poses great difficulties for the check airman conducting a LOFT exercise. On the one hand, he or she must contend with the fact that there are neither validated measures available to use in assessing crew process and performance (other than measures of technical flying skill), nor any single best way to conduct a flight safely and competently--matters we have discussed previously. But beyond those problems, check airmen experience great difficulty in balancing the training and evaluation components of LOFT exercises. They are, for example, extremely reluctant to give "unsatisfactory" ratings for LOFT, using the argument that "if the crew found it a significant learning experience, it was a satisfactory session regardless of the performance exhibited." On the other hand, check airmen are deeply troubled by the prospect of releasing for continued line flying pilots whose

---

[8] Consider, for example, an individual who is competent in all technical flying skills and functioning well as a co-pilot--but whose capacity to fill a captain's role is questionable. A scenario could be constructed to allow that individual to demonstrate his or her decision-making and managerial skills by serving as a captain on a simulated flight.

behavior in the exercise revealed serious safety-related problems.

In our view, the LOFT technology provides an opportunity to provide air transport with an excellent means of pursuing both training and performance evaluation objectives. But this opportunity will be realized only if several developments occur. First, as noted earlier, is the development of an assessment technology that is accepted by operational personnel as being reliable, valid, and objective. Second is achievement of a reduction in the pressures against evaluation operating on both airline management and pilot groups. And third is the development of a means of using LOFT that threads a course between the two horns of the training-evaluation dilemma.[9]

Research Approaches

The objective of our research is to generate means of understanding, measuring, and constructively influencing team performance--and to do so in ways that promote both improved organizational practice and the accumulation of scholarly knowledge about groups and group effectiveness. Although this paper is the first joint research or writing we have done, our interests have been converging in recent years as both of us have experienced the engagement and the frustration of trying to make sense of groups and to figure out what might be done to help them perform more effectively.

Helmreich has been mainly concerned with the isolation of personality and motivational factors relevant to individual and group performance, especially as they relate to flightcrews (Helmreich & Spence, 1978; Helmreich, 1982;

---

[9] Our ideas about how this might be done, which are still under development, are described in a companion paper (Helmreich & Hackman, 1984). In brief, we propose a means of partitioning analyses of individual and crew performance in LOFT exercises, and we suggest development of a second version of the technology (called LOCK, for Line Oriented Check) intended explicitly for use in formal assessments.

1983). He also has examined the effects on performance of composing crews with differing personality constellations and the ability of various training procedures to counter or enhance the behavioral effects of personality. Hackman (e.g., 1982; 1983) has focussed his recent research on task and organizational factors that affect group processes and group task effectiveness. He has developed a normative model that specifies aspects of teams and situations that may be particularly potent in promoting excellent performance, and that organizes those factors in a way that invites their use in the design and management of task-performing teams. In collaboration with Robert Ginnett, he is currently in the process of revising the normative model for specific application to cockpit crews.

In the sections that follow, we sketch some of the major features of these two research programs. As will be seen, both programs seek better ways of conceptualizing and assessing cockpit crew processes, with Helmreich approaching the problem from his research on individual differences, and Hackman from his research on task and organizational variables. Both programs are committed to the development of a descriptive empirical database against which theoretical constructs can be tested and the impact of interventions assessed.

The Helmreich Project

This approach to the assessment of team processes and performance is explicitly multidimensional, including observations and ratings both in unconstrained line operations and in controlled flight simulations that present the same operational problems to a number of crews. In addition to observer judgments, self-assessments by crewmembers following simulator flights are collected to understand participants' perspectives on the processes and outcomes of flight segments.

An important element of the approach involves the development of multiple coding schemata designed to capture the molecular aspects of performance enactment. Coding categories are evaluated using time-lined videotapes of a LOFT scenario flown by line crews. Three broad areas are specified: information transfer, control, and group climate. Information transfer components include both operational and social-emotional communications, as well as breakdowns of the relative contributions (initiated and reactive) of team members, and the qualitative aspects of the interaction (i.e. the forms of communication). Control factors consist of direct and indirect attempts to influence and "manage" the ongoing situation. Climate refers to indicators of the affective tone of group interactions and the inferred states of individual team members. No attempt has been made to impose independence on the behavioral categories; they are related cuts of the same phenomena.

Process variables such as those just described are difficult to interpret except within the context of the task situation. For this reason, several different frames of reference are being explored. The most basic consists of examining each phase of flight (pre-flight, take-off, climb, cruise, descent, approach, and landing) discretely and, within each phase, classifying the situation as normal, acute non-standard, or continuing non-standard. Another approach involves classifying activities in terms of their relationship to necessary actions during each phase of flight. That is, actions may be directed towards coping with the immediate situation, may be attempts to complete activities that should have been accomplished earlier but were deferred, or may be focussed on future actions and the development of action strategies. A final approach consists of utilizing captain behavior as a benchmark against which to measure the behaviors of the other team members.

At this stage in the research, it is impossible to tell how useful each of these approaches may be, or if some combination of measures and referents will prove most informative.

After preliminary evaluation of alternative behavioral coding strategies, other phases of the research will involve composing crews on the basis of personality and demographic characteristics and exposing them to the same LOFT scenario. Additional research questions involve assessment of the effects on group process and performance outcomes of different training techniques, especially training in crew coordination and cockpit resource management. A particularly important applied objective of the research is the development of relatively simple evaluation categories that can be used by operational personnel to expand and improve the formal evaluation process.

The Hackman-Ginnett Project

The normative model on which this project is based posits that the overall effectiveness of a work team is a joint function of three factors:

  --the level of effort group members collectively expend carrying out task
    work,

  --the amount of knowledge and skill members bring to bear on the group
    task, and

  --the appropriateness to the task of the performance strategies used by
    the group in its work.

We refer to effort, knowledge and skill, and performance strategies as process criteria of effectiveness. They are the hurdles a group must surmount to be effective. To assess the adequacy of a group's task processes, then, we might ask: Is the group alert enough and working hard enough to get the task done well and on time? Do members have the expertise required to accomplish the task, and are they using their collective knowledge and skill efficiently?

Has the group developed an approach to the work that is fully appropriate for the task being performed, and are members implementing that strategy well? Answers to these questions provide useful diagnostic data about a group's strengths and weaknesses as a performing unit, and they are the conceptual hook on which the rest of the research hangs.

Three classes of variables are specified as particularly good points of leverage for creating conditions that foster achievement of the process criteria: (a) how the group is designed (including properties of the team task, the composition of the team, and the core norms that regulate member behavior); (b) the level of support it receives from the organization (with special attention to the adequacy of material resources needed by the team, and to organizational reward, education, and information systems); and (c) how the role of the group leader (or manager) is structured and the behavior of the person who occupies that role (with with special attention to condition creating, team building, and process management activities).

A set of instruments is under development to assess both the criterion measures and each of the condition-setting variables as they apply specifically to cockpit teams. These measures will involve the use of multiple methodologies whenever possible to triangulate on the concepts being assessed. Survey and interview methods will be used to assess the chronic state of variables that are not expected to vary substantially in the short term (e.g., aspects of the organizational context), and to obtain crewmembers' perceptions of their team and its work. Intense, detailed observations and descriptions of crew behavior will be collected at "task critical" and "group critical" times in the life of the group. (These are specifiable occasions when what happens next is likely to significantly affect the group's

performance or its viability as a performing unit, respectively.) Critical incident techniques will be used to capture significant events that occur at unpredictable times.

Based on what is learned from data collection activities (including both cockpit observations and studies done in simulators), the measures will be revised and retested until (a) they are usable by a trained observer/interviewer without excessive difficulty, and (b) they can be shown to capture gross differences on variables of research interest. At that point, a more systematic set of research activities will be instituted, to validate the instruments and to assess their usefulness in training and evaluating cockpit crew members.

The findings from the Helmreich and Hackman-Ginnett research programs will be integrated and evaluated using specially-designed LOFT scenarios. Data from these exercises will be used to develop a parsimonious hybrid assessment system that builds on the common and unique features of the two research programs. The hope is that the hybrid system will prove useful both as a research tool and, in abbreviated form, as a reliable technology for assessment in both operational and crew training environments.

Conclusion

We began this writing project in hopes of surfacing some general issues and insights about the assessment of teams that do work in organizations. Yet virtually the entire paper has been devoted to exploration of the special challenges faced in attempting to assess the behavior and performance of crews that fly aircraft for commercial airlines. Have we slipped off the mark, and written a paper that will be of interest only to a very small group of researchers with special interest in cockpit crews?

That is, of course, for the reader to decide. Our belief (and certainly our hope) is that even readers with no interest in cockpit crews will find here some issues that also are salient in assessing other kinds of task-performing groups and teams. Are there teams for which historical, political, and organizational contexts do not significantly constrain and direct assessment activities? Does any team generate objective performance outcomes that everyone agrees capture precisely how well the team has functioned? Are there any managers who are untroubled about their need to rely on subjective judgments about group processes, or any team members who do not worry about those judgments being used capriciously or unfairly? Are the internal processes of any team unaffected by organizational structures and systems, matters over which team members may have little control--but that can strongly affect how (and how well) members work together? Do we know of any team for which the tension between training/development and assessment/control is not a serious problem, or any organization that does not have difficulty using constructively the rich informal assessments that exist about teams and the contributions of their members?

The challenges in assessing task-performing teams, we believe, are as pervasive as they are difficult. We hope that by writing about how those challenges are manifested in cockpit crews we may have provided at least a few ideas or leads that will be useful to other researchers concerned with the assessment of other teams in other contexts.

References

Cooper, G. E., White, M. D. & Lauber, J. K. (Eds.), Resource management on the flight deck (NASA Report No. CP-2120). Moffett Field, CA: NASA Ames Research Center, 1979. (NTIS N80-22283)

Gersick, C. J. G. Life cycles of ad hoc task groups (T. R. No. 4). New Haven, CT: Research Program on Group Effectiveness, Yale School of Organization and Management, 1983.

Gersick, C. J. G. The life cycles of ad hoc task groups: Time, transitions, and learning in teams. Unpublished doctoral dissertation, Yale University, 1984.

Hackman, J. R. A set of methods for research on teams (T. R. No. 1). New Haven, CT: Research Program on Group Effectiveness, Yale School of Organization and Management, 1982.

Hackman, J. R. A normative model of work team effectiveness (T. R. No. 2). New Haven, CT: Research Program on Group Effectiveness, Yale School of Organization and Management, 1983.

Hackman, J. R. The design of work teams. In J. W. Lorsch (Ed.), Handbook of organizational behavior. Englewood Cliffs, NJ: Prentice-Hall, in press.

Helmreich, R. L. Pilot selection and training. Paper presented at the American Psychological Association annual meeting, Washington, D.C., 1982.

Helmreich, R. L. What changes and what endures: The capabilities and limitations of training and selection. In N. Johnston (Ed.) Proceedings of the Aer Lingus/Irish Airline Pilots Association Flight Symposium, Dublin, Ireland, 1983.

Helmreich, R. L., Psychological issues in space station planning and design. Unpublished manuscript. Austin, TX: University of Texas, 1984.

Helmreich, R. L., Cockpit management attitudes. Human Factors, in press.

Helmreich, R. L. & Hackman, J. R. Evaluating flightcrew performance: Policy, pressures, pitfalls, and promise. Unpublished manuscript. Austin, TX: University of Texas, 1984.

Helmreich, R. L. & Spence, J. T. The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. JSAS Catalog of Selected Documents in Psychology, 8(35), Ms. 1677, 1978.

Katz, D. & Kahn, R. L. The social psychology of organizations (2nd ed.). New York: Wiley, 1978.

Lauber, J. K. & Foushee, H. C. Guidelines for line oriented flight training (Vol. 1) (NASA Report No. CP-2184). Moffett Field, CA: NASA Ames Research Center, 1981.

Mearns, D.J. FDR--The pilot's friend: BA/BALPA co-operation in action. In
     N. Johnston (Ed.) Proceedings of the Aer Lingus/Irish Airline Pilots
     Association Flight Symposium, Dublin, Ireland, 1983.

Melton, A.W. Army air forces (Aviation Psychology Program Report No. 4).
     Washington, DC: Defense Documentation Center, 1947.

Nance, J. J. Splash of colors: The self-destruction of Braniff
     International. New York: Morrow, 1984.

National Transportation Safety Board. Aircraft accident report (NTSB Report
     No. AAR-79-7). Washington, D.C.: NTSB Bureau of Accident Investigation,
     1979.

National Transportation Safety Board. Aircraft accident report (NTSB Report
     No. AAR-81-11). Washington, D.C.: NTSB Bureau of Accident Investigation,
     1981.

National Transportation Safety Board. Aircraft accident report (NTSB Report
     No. AAR-82-8). Washington, D.C.: NTSB Bureau of Accident Investigation,
     1982.

Porter, L. W., Lawler, E. E. & Hackman, J. R. Behavior in organizations. New
     York: McGraw-Hill, 1975.

Tyler, L. E. Thinking creatively. San Francisco: Jossey-Bass, 1983.